DOCUMENT RESUME

ED 162 559 HE 010 578

AUTHOR Darr, Ralph F., Jr.
TITLE Evaluation of College Teaching: State of the Art, 1977.
PUB DATE Apr 77
NOTE 25p.; Paper presented to the Ohio Academy of Science (Capital University, Columbus, Ohio, April 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; Administrative Personnel; Bibliographies; *College Faculty; Evaluation Criteria; *Evaluation Methods; *Faculty Evaluation; *Formative Evaluation; Higher Education; *Peer Evaluation; Rating Scales; Self Evaluation; State of the Art Reviews; *Student Evaluation of Teacher Performance; Summative Evaluation

ABSTRACT

Student ratings of teacher performance in higher education have received the bulk of research interest and effort since the first formal published rating form in 1926. Peer ratings run a poor second, and little has been done in the areas of self-evaluation and administrative evaluation. Attempts at using student achievement as an evaluation criterion have run into their own unique difficulties. This state-of-the-art review summarizes sources of evaluative data (including student ratings, peer ratings, self-evaluation, and administrator evaluation), and analyzes faculty evaluation systems. It is contended that the future holds two basic trends for evaluation of instruction: (1) more systematic evaluation and (2) more attention to the theoretical basis of evaluation. A five-page bibliography is included. (Author/MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED162559

EVALUATION OF COLLEGE TEACHING:

STATE OF THE ART, 1977

Ralph F. Darr, Jr.

The University of Akron

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Ralph F. Darr Jr.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

HE 010 578

This paper was presented at The Ohio Academy of Science Psychology Division; Capital University, Columbus, Ohio, April 1977.

EVALUATION OF COLLEGE TEACHING:

STATE OF THE ART, 1977

After thirty years of studying evaluation of higher education, H. H. Remmer (1958) stated,

> No teacher has any choice as to whether he wishes to be judged by his students. The only choice he has is whether he wishes to know how he is judged and possibly capitalize on his feedback.

It is the premise of this paper that Remmer's proposition holds true today, almost twenty years later, and it is true for peer, self and administrative evaluation as well. Further, it is our feeling that users of these instruments should know what factors influence ratings and those which do not. The first part of the paper will look at the state of the art in 1976-1977. In the second section, we will attempt a glance into the future.

STATE OF THE ART, 1977

Sources of Evaluative Data

Student ratings have received the bulk of research interest and effort since the first formal published form - the Purdue Rating Scale of Instruction (1926). Peer ratings run a poor second. Little has been done in the area of self-evaluation and administrative evaluation. Attempts

at using student achievement as a criterion for evaluating instruction have run into their own unique difficulties.

Student Ratings

A variety of factors have been studied to assess their influence on student ratings of instructor and instruction. These factors can be grouped into four clusters:

1. student characteristics
2. faculty characteristics
3. environment or organization characteristics
4. characteristics of the instruments themselves.

Student characteristics. Recent studies of the relationship between expected grades (Velicher, 1977, and Bejar and Doyle, 1976) and student ratings of instructors and courses were found to be nonsignificant as was the relationship between grade received and student ratings (Brown, 1977).

Centra (1977), however, found student's global ratings of instruction to be highly related to mean examination performance. Peck (1977) found that the degree to which the student was able to meet course requirements in the time allotted, received the intellectual stimulation expected in the course, and the degree to which the student performed up to his or her expectations in the course influenced his or her evaluation of instruction.

Bejar and Doyle (1976) found student ratings of instruction to be uninfluenced by sex of the student or level of the course.

Haslett (1976) and Shikiar (1976) found students to be consistent in their selection of the characteristics of effective or good teachers. Haslett's students suggested good teachers demonstrate: a.) good teacher-student rapport, b.) appropriate communicative style, c.) effective instructional style, and d.) stimulation. Shikiar subjects' data produced two dimension: a.) personal vs. professional characteristics, and b.) learning distraction characteristics vs. learning aids.

Villano (1977) and Butler and Tipton (1976) found students to be consistent over time in their rating of instruction and instructors. Abrami, et al. (1977) suggest that further research is required in order to determine just what effect student characteristics have on student ratings of instructors and instruction. Ascher (1977) attempted to anchor student ratings by having the students first identify the characteristics of the effective teacher and then have the students evaluate their current instructor on those characteristics. The individual teacher's rating would be in terms of variance from the ideal. In this way, Ascher attempted to measure student bias going into evaluation. Significant trends in bias were not isolated.

Recent studies indicate that such student variables as sex, grade level, age, and major are not significantly related to student ratings of the instructor or the course.

Current research seems to be focusing on the cognitive and affective (Ascher, 1977 and Decton, et al., 1977) predisposition and its influence on student evaluation of instruction.

Faculty characteristics. Generally, there has been less research that has looked at the characteristics of faculty who receive high and low ratings as instructors or their course. Part of the difficulty is methodological, in that such studies would require comparisons across courses and disciplines.

Gillmore and Kane (1977) found that at least five classes for each instructor had to be evaluated by different students before dependable results for generalizing over courses for the instructor could be obtained. No dependable results were obtained for course when attempting to generalize across instructors, thus suggesting that the instructor accounts for more of the variance in instruction than course content, per se.

Dudek, at al., (1977) are approaching the study of faculty characteristics by first seeking agreement among raters as to what college teachers do. Spearman-Brown reliabilities range from .678 to .948 (average .841). Four clusters have been isolated. The team is in the process of developing instruments that focus on the clusters of faculty characteristics. Dixon and Koerner (1976), working with nursing students and faculty, were able to isolate the perceived characteristics of effective teaching faculty;

however, their attempts to develop an instrument for students to evaluate those characteristics has been unsuccessfull to date.

Delaney and Kojaku (1977) investigated the relationship teaching experience of faculty, their participation in faculty development activities and student ratings of instruction. Controlling for students expected grade and class size, the overall length of teaching experience was found to have considerable positive effect on the variance in ratings. Preservice training was unrelated to ratings, while inservice training was only mildly related to evaluation results. Morrison and Friedman (1977) found "faculty socialization experience" and "educational orientation" to be unrelated to student ratings.

In an experimental study, Perry (1977) was able to demonstrate that student ratings of faculty could be manipulated by the information given students prior to attending a lecture. Half of the students received positive evaluation of the lecturer. Half received negative evaluations. The lecture styles were varied, two dimensional: 1.) high/low content, and 2.) high/low teaching style. Student ratings tended to move in the direction of the pre-lecture information rather than either lecture style or content.

Butler and Tipton (1976) found that teaching behavior can change as a result of student feedback, and students can recognize those changes.

The amount of research on teacher characteristics is minimal. Perry seems to be a step in the right direction.

Environmental or organizational characteristics. Recent research in the area of organizational characteristics has been minimal. It never was extensive, partly because there is not much that the instructor or the student can do to change them. Neither the instructor, nor the student has much to say about class size, level of the course in the sequence, time of day or room location.

There is still considerable argument about the influence of class size and course level. Sheehan (1975 and 1976) argues that class size is negatively related to student ratings; Aleamoni (1976) argues that they are unrelated, and Overall (1977) reports class size to be related to student ratings in a curviliniar fashion with small and large class receiving the higher ratings.

Pohlman (1977) reported agreement among the faculty from five disciplines as to the five most important characteristics of effective teachers: a.) knowing when student understands, b.) increasing student appreciation of the subject matter, c.) answering impromptu questions, d.) achieving the objectives of the course, and e.) giving several examples to explain complex topic. With appropriate items to sample each of the five areas adequately so that reliability might be established, it should be possible to generate instrument that cuts across disciplines and other artificial divisions in colleges and

universities.

Reed (1977) found the Kansas State University's Instructional Development and Effectiveness Assessment (IDEA) instrument was indeed influenced by factors extraneous to the instrument.

Clearly, additional effort must be devoted to the investigation of organizational pattern. With dropping enrollment and increasing concern about teaching and maintaining enrollment, possibly something can be done about these previously unmodifyable variables like class size if more data can be obtained about their effect on the teaching/learning process.

Peer Ratings

Peer evaluation is at the very heart of university life, yet very little has been published about the procedures, rationale and criteria faculty members employ in making decisions that affect their colleagues.

History. The American Association of University Professors has been the spokesman for teaching faculty and as such has on occasion offered recommendations on the issue of faculty evaluation.

In 1939, the University of Michigan Chapter of the AAUP recommended that teaching ability "be judged by visitation to classes by colleagues whose individual judgments are to be combined to give a composite opinion,

and by students through the proposed questionnaire." Smart (1965) noted that AAUP had neglected to formulate a committee to deal with evaluation specifically, but rather had left the decision to administration and the faculty of the various institutions. Peer evaluation was favored over administrative.

Current State of the Art. There are few isolated attempts to develop and validate peer rating scales. Nadeau (1977) suggests that peers be asked to rate colleagues on the same instrument that students use and that the data from students' ratings and colleagues' evaluations be compared. J. O. Miller (1977) has developed a high inference peer evaluation instrument which is congruent with an instrument employed by students. Ballard, et al., (1976) obtained high agreement between student and faculty-peer ratings of teaching performance ($r = .84$) and instructor "likeableness" ($r = .86$).

Factors Related to Peer Evaluations. The findings of the Ballard (1976) study and others have led some authorities to suggest that high correlation between student and peer rating negate the need for peer evaluation. Currently, little is reported about factors that relate to high and low ratings of colleagues as teachers, quite likely because there is limited instrumentation and data from peer evaluations. However, two recent studies attest to the depth of study surrounding peer evaluations. Wood (1977)

reported that peers' ratings of colleagues' teaching effectiveness was positively correlated with office location, i.e., faculty located in main pathways received higher ratings than those located in more remote areas. Colleague ratings on research and service were not related to room location. Student ratings of instruction and instructor were found to be more reliable than colleague ratings. von Rosenstiel and Schuler (1975) found peer evaluation to be significantly related to the first letter of the colleagues last name.

Self-Evaluation

If little substantial is written about peer evaluation even less is written about self-evaluation. In Western society, we seem to distrust our own inner feelings or at the least not be willing to present them or let others present them as admissible evidence of thought or work. Yet evidence presented by the planner, executor and evaluator does seem to have a part to play.

In public education, there is renewed interest in self-evaluation (Simpson, 1966). There are a few checklists in higher education, but they are not receiving much current attention. Nadeau (1977) suggests that the instructor might start self-evaluation by taking the same course evaluation as his student and make an effort to answer each item objectively from his point of view as instructor. The instructor should evaluate his responses by comparing

them with the responses of his students, extremes as well as central tendencies.

Administrative Evaluation

Historically, administrators, deans and above, have not been directly involved on a day-by-day basis with the evaluation of teaching, except in a few small institutions. In most cases, deans because of time schedules, and other more pressing concerns, tend to make their decisions about instructors and courses on the basis of data supplied to them by others: students, colleagues, peers, department heads, and/or superiors (Smock and Crooks, 1973).

Recent concern over dropping enrollment and financial difficulties have forced many deans and other upper-level administrators to renew their interest in evaluation of teaching (Seldin and Watkins, 1974). Deans will probably continue to use data gathered by others but they will widen the data base. Seldin and Watkins suggest that deans in small colleges want all the evidence that they can get that a particular instructor's teaching draws and maintains students before they commit themselves to tenure, promotion or salary increases. There is no money on the small campus for the faculty members whose major successes are outside the classroom, whether they are in the laboratory or the consulting arena.

The University of Illinois' COPE program is an attempt to get information about teaching effectiveness, research and service on the basis of departmental evaluations into the hands of deans and upper-level administrators in a systematic fashion (McIntyre, 1977).

It is very unlikely that in all but a few very small schools that upper-level administrators will become involved in direct evaluation of instruction.

Achievement of Students as Criterion of Teaching Effectiveness

At the 1977 Annual Meeting of the American Educational Research Association, the Special Interest Group, Instructional Evaluation, rejected the achievement of students as a criteria for evaluating the effectiveness of instruction and the instructor, not so much that it rejected the notion that producers should be evaluated on the basis of his or her products as that it was hard pressed to identify the products and how to measure them (Aleamoni, 1977). What shall be the criterion of achievement? a test score? job placement? acquisition of competences? national or international renown in twenty years? If there is no agreement of the criterion or criteria, then prediction is impossible. Task analysis and specification of competencies students should achieve may be the answer, but at this point, they appear to be a long way off (Coker, 1977, and Sanders, 1977).

Faculty Evaluation Systems

Historically, the university and its faculty have had three major roles: 1.) teaching, 2.) research, and 3.) service. In 1939, the University of Michigan Chapter of the AAUP extended the three areas to six: 1.) teaching, 2.) research, 3.) standing in the profession, 4.) personal qualities, 5.) value in departmental and university administration, and 6.) value of public and community service. Many institutions in the years to follow adopted similar resolutions with some slight modification to take care of homegrown biases. Almost all put teaching first on the list and many assigned it high weight, e.g., 60 percent of any evaluation, then promptly found ways to get around having to evaluate teaching. Some faculty rejected student evaluations as violating their academic freedom. Even more responded that observing colleagues in their classroom was infringement of privacy and academic freedom. Still other institutions handled the matter by straight forward acknowledging that all of their teaching faculty were great teachers lest they would not be at that great institution; then promptly took care of teaching problems by shuttling less adequate teachers off to administration, research or branch campus, where they were out of the department's way and their bad marks fell on someone else's head. Publications and service could be handled by simple tallying. No evaluating needed.

More recently, the declining enrollments, decreasing funds and faculty surplus have brought new emphasis to faculty evaluation. Faculty have to earn their way. With revenue from research declining, the student becomes a more important source of revenue. He or she must be wooed to the school and maintained there. The instructor becomes the key link in the process.

Faculty evaluation and development is becoming an increasingly important part of the operations of many larger institutions, e.g., COPE at the University of Illinois, Teaching Resource Center at the University of California (Davis), Measurement and Research Center at Purdue, etc. Personnel working in faculty evaluation and development at each of the institutions are devoting a large portion of their time to the evaluation of instruction as well as research and service (Seibert, 1977 and Dudek, 1977).

THE FUTURE

The future holds two basic trends for evaluation of instruction: 1.) more systematic evaluation, and 2.) more attention to the theoretical basis of evaluation. These trends will be dictated by rising costs, decreasing enrollment, decreasing funds and greater emphasis upon accountability for the use of public funds.

Systematic Evaluation

Systematized evaluation will replace the haphazard procedures of the last fifty years. Hopefully, evaluation of teaching will be a prominent part of the system. Possibly, those who do the teaching will get some recognition for their efforts.

Sources of Evaluative Data

Each of the five sources of data will continue to play a role in the system, but hopefully, each role will be unique.

Student ratings. Hopefully, the number of instruments in use will lessen with the development of instruments that have the built-in flexibility of IDEA and CAFETERIA to replace the old, ridged global questionnaires.

Secondly, greater emphasis should be placed on the contribution and limitations of student ratings. The student knows best how he or she feels about what went on during lecture, or an exam or a film. On the other hand, very few undergraduates are sufficiently skilled to make decisions about appropriateness of curriculum.

Ideally, data from student ratings will begin to influence educational practice (Butler and Tipton, 1976). Most items will be diagnostic in nature. Only a few will be summative and they will form a core across all administrations of the instrument. To be of value to the instructor, items must either assess characteristics of

the instructor or the characteristics of the course. Items will be brief, clear and addressed to specific aspects of the teaching/learning process.

Peer ratings. Initially, the number of instruments and amount of research will have to increase markedly. Eventually, however, a few efficient, reliable and valid instruments should emerge. The correlates of high and low ratings will be teased out.

The unique contribution of peer-ratings will become more widely recognized. Peers are in the best position to assist with decision about curriculum, e.g. what should be taught? what methodology employed? what evaluative techniques are most appropriate? Certainly, students with limited background cannot help. The department heads and deans have their own troubles. Traditionally, the faculty member has had to go it alone or isolate one or two colleagues who might or might not share his concerns. Formal evaluations by colleagues would extend th individual's peer base.

With the current committee structure of most universities, colleagues will be called upon to play both formative and summative roles. Initially, they might suggest to the neophyte how he might improve his instruction - a formative function. Later in a summative role, they might assess how well he or she treated their suggestions. The role of peer input into the personnel

decision making process will become clear. Faculty will be evaluated for tenure, promotion and salary on the basis of demonstrated behavior and improvement rather than hearsay.

Self-rating. Like peer ratings, a great deal of work is necessary in this area. Checklists and self-rating scales must be developed. Ease of use should receive maximum concern. These instruments should help the instructor get a clearer picture of how he or she spends time in planning, executing and evaluating instruction and whether the investment is worthwhile. Only the individual instructors know how much time he or she spent planning or how he or she felt after the lesson was presented. Self-ratings provide information to the instructor and other evaluators that can be obtained from no other source; however, for most situations, self-report data is best used in formative ways in academea.

Administrative Evaluation. Deans and above in most larger institutions will continue to resist efforts to draw them directly into evaluation of teaching. They will continue to press for clear, precise summative data from an increasingly wider set of sources: students, departmental peers, department heads, etc. They will primarily leave formative evaluation to department heads and colleagues.

Achievement of students. Current emphasis upon accountability, matching students to jobs, task analysis and competency-based education are putting pressure on measure-

ment experts to develop adequate criteria for evaluating learning outcomes by means other than our current paper and pencil tests. Greater effort will be put into task analysis and specification of necessary competencies to accomplish the task. Traditional, assumptions about competencies necessary to perform certain tasks have broken down (Coker, 1977).

The Role of Theory

More attention will be directed to both the theory of instruction and the theory of evaluation underlying the various instruments.

Theory of Instruction. Instruments will have to be develped that are appropriate for a variety of instructional procedures. Increasingly, instructors are employing a variety of instructional procedures. The traditional instrument based on the teacher as teller style of instruction is often totally inappropriate for experiential courses. IDEA and CAFETERIA are steps in this direction.

Theory of Evaluation. The purpose of persons who construct instruments for evaluating teaching is shifting and will continue to shift from getting neat data to standardize instruments that can be published and sold elsewhere to generating instruments that help the individual faculty, department heads, and departmental committees make decisions about the teaching/learning process. The

emphasis is from summative toward formative.

From the teaching faculty's point of view, we are headed in the right direction.

BIBLIOGRAPHY

Abrami, Philip C.; Leventhal, Les. and Perry, Raymond R. The generalizability of student ratings of instruction. American Educational Research Association, 1977 Annual Meeting, Abstract of Papers and Symposia. New York: April 4-8, 1977, p. 101.

Aleamoni, Lawrence M. Comments on "On the the Invalidity of student ratings for administrative personnel decisions." Journal of Higher Education, 1976 (Sept. - Oct.), Vol. 47(5), 607-610.

Aleamoni, Lawrence M. (Organizer) and McKeachie, Wilbert J. (Chair). Instructional evaluation in rank, pay, and tenure decisions (SIG/ Evaluation of Instruction, Symposium). American Educational Research Association, 1977 Annual Meeting, Abstract of Papers and Symposia. New York: April 4-8, 1977, p. 124.

Ascher, Gordon. Students' evaluation of teachers as a function of student ethnicity and personal values. American Educational Research Association, 1977 Annual Meeting, Abstract of Papers and Symposia. New York: April 4-8, 1977, p. 111.

Ballard, Michael; Rearden, J. and Nelson, Linda. Student and peer rating of faculty. Teaching of Psychology, 1976 (Apr.), Vol. 3(2), 88-90.

Bejar, Isaac I. and Doyle, Kenneth O. The effect of prior expectations on the structure of student ratings of instruction. Journal of Educational Measurement, 1977(Summer), Vol. 13(2), 151-155.

Brown, David Lile. Faculty ratings and student grades: A university-wide multiple regression analysis. Journal of Educational Psychology, 1976 (oct.), Vol. 68(5), 573-578.

Butler, Joel R. and Tiption, Robert M. Rating stability shown after feedback of prior ratings of instruction. Improving College and University Teaching, 1976 (Spr.) Vol. 24(2), 111, 112, 115.

Centra, John A. Student ratings of instruction and their relationship to student learning. American Educational Research Journal, 1977(Winter), Vol. 14(1), 17-24.

Coker, Homer (Chairperson), Process/product research in the real world: An attempt to develop and validate a competencey based teacher certification system. American Educational Research Association, 1977 Annual Meeting, Abstracts of Papers and Symposia. New York: April 4-8, 1977, p. 112.

Deaton, William L; Peggio, John P. and Glasnapp, Douglas R. A scale to assess the affective entry level of students. 1977 NCME Annual Meeting, Paper and Symposia Abstracts. New York: April 5-7, 1977, p. 11.

Delaney, Edward L. and Kojakee, Lawrence K. The influence of teaching experience and faculty development activities on student ratings of instruction. American Educational Research Association, 1977 Annual Meeting, Abstracts of Papers and Symposia. New York: April 4-8, 1977, pp. 167-168.

Dixon, Jane K. and Koerner, Beverly. Faculty and student perceptions of effective classroom teaching in nursing. Nursing Research, 1976 (July - August), Vol. 25(4), 300-305.

Dudek, Judith Love; Minnis, Douglas L.; Sandoval, Jonathan and Yonge, George D. An instrument to describe college teaching: A reliability study. American Educational Research Association, 1977 Annual Meeting, Abstracts of Papers and Symposia. New York: April 4-8, 1977, p. 60.

Gillmore, Gerald M. and Kane, Michael T. The instructor and the course as units of analysis in student instructional ratings. 1977 NCME Annual Meeting, Paper and Symposia Abstracts. New York: April 5-7, 1977, p. 5.

Haslett, Betty J. Dimensions of teaching effectiveness: A student perspective. Journal of Experimental Education, 1976 (Summer), Vol. 44(4), 4-10.

McIntyre, Charles J. COPE: A systematic approach to the evaluation of academic departments. in Seibert, Warren F. (Organizer), Symposium: Teaching, academic programs, and ratings: To improve or to judge? AERA 1977, Annual Meeting, Abstract of Papers and Symposia. New York: April 4-8, 1977, p. 93.

Miller, J. O. "The construction and validation of selected high inference measures for post-secondary teacher evaluations." in Weber, Margaret (Organizer) Symposium: The Methodological implications of the use of high inference variables for judging teacher effectiveness. American Educational Research Association 1977, Annual Meeting, Abstract of Papers and Symposia. New York: April 4-8, 1977, p. 125.

Morey, T. Mark; Garthwaite, Roger A. and Zimowski, Michele F. Ideal, actual and importance ratings of teaching effectiveness: A search for validity. Program: 49th Annual Meeting of the Midwestern Psychological Association. Chicago, Ill.: May 5-7, 1977.

Morrison, James L. and Frieman, Charles P. A causal analysis relating faculty socialization experience and educational orientations to perceived teaching effectiveness. American Educational Research Association, 1977 Annual Meeting, Abstracts of Papers and Symposia. New York: April 4-8, 1977, p. 168.

Nadeau, Gilles G. A plan for evaluation for promotion, tenure and assignment in Aleamonie, L. M. (Organizer), Instructional evaluation in rank, pay and tenure decisions (SIG/Evaluation of Instruction Symposium). American Educational Research Association, 1977 Annual Meeting, Abstracts of Papers and Symposia. New York: April 4-8, 1977, p. 124.

Overall, J. U. Class size and student ratings of instruction: A clarification of relationship. American Educational Research Association, 1977 Annual Meeting, Abstracts of Papers and Symposia. New York: April 4-8, 1977, p. 168.

Peck, Teresa Harris. The prediction of teacher evaluation scores from knowledge of student variables. 1977 Annual Meeting of the American Educational Research Association, Abstracts of Papers and Symposia. New York: April 4-8, 1977, p. 59.

Perry, Raymond. Educational seduction: The effect of teacher reputation on student satisfaction and learning. American Educational Research Association, 1977 Annual Meeting, Abstracts of Papers and Symposia. New York: April 4-8, 1977, p. 149.

Pohlman, John T. A description of effective college teaching in five disciplines as measured by student ratings. Research in Higher Education, 1976, Vol. 4(4), 335-346.

Reed, Jeffrey G. Effect of miscellaneous and uncontrolled variables on student evaluation of instructors and courses. Program: 49th Annual Meeting of the Midwestern Psychological Association. Chicago, Ill.: May 5-7, 1977.

Remmers, H. H. On student perceptions of teacher's effectiveness. In McKeachie, Wilbur J. The Appraisal of Teaching in Large Universities. Ann Arbor, Michigan, 1958.

Remmers, H. H. and Brandenburg, G. C. Experimental data on the Purdue Rating Scale for Instruction. Educational Administration and Supervision. 1927, Vol. 13, 519-527.

Report of a Committee of The University of Michigan Chapter of the AAUP. The Evaluation of Faculty Services. Ann Arbor, Michigan: University of Michgan Administrative Services, Vol. 1, No. 3, 1939.

Sanders, James R. (Moderator). Test development issues in competency based measurement. 1977 NCME Meeting, Paper and Symposium Abstracts. New York: April 4-8, 1977, p. 16.

Seibert, Warren F. (Organizer). Teaching, academic programs, and ratings: To improve or to judge? American Educational Research Association, 1977 Annual Meeting, Abstract of Papers and Symposia. New York: April 4-8, 1977, p. 93.

Seldin, Peter and Waking, Edward. How deans evaluate teaching. Change, 1974 (Nov.), Vol. 6(9), 48-49.

Sheehan, Daniel. On the invalidity of student ratings for administrative personnel decisions. Journal of Higher Education, 1975 (Nov.-Dec.), Vol. 46(6), 687-700.

Sheehan, Daniel S. Reply to L. M. Aleamoni. Journal of Higher Education, 1976 (Sept.-Oct.), Vol. 47(5), 610-611.

Shikiar, Richard. Student and faculty perceptions of teacher characteristics. Journal of Psychology, 1976 (Mar.), Vol. 92(2), 215-218.

Simpson, Ray H. Teacher self-evaluation. New York: Macmillan. 1966.

Smart, R. C. The Evaluation of Teaching Performance from the Point of View of the Teaching. Chicago: American Psychological Association Meeting, 1965.

Smock, H. Richard and Crooks, Terence J. A plan for the comprehensive evaluation of college teaching. Journal of Higher Education, 1973 (Nov.), Vol. 44(8), 577-586.

Velicher, Wayne. Expected grade and student ratings of instructors. A reappraisal. Abstracts of Papers and Symposia: 1977 Annual Meeting of the American Educational Research Association. New York: April 4-8, 1977, p. 59.

Villano, Maurice W. The developmental and validation of an omnibus student rating form for evaluation of courses and instructors at a large university. 1977 NCME Annual Meeting, Paper and Symposia Abstracts. New York: April 5-7, 1977, p. 3.

von Rosenstiel, Lutz and Schuler, Heinz. [A for Arnold, B for Bender: On the sociodynamics of the academic career.] (Germ.) Psychologische Rundschau, 1975 (Jul.), Vol. 26(3), 183-190.

Wood, Peter H. The description and evaluation of college department's faculty rating system. American Educational Research Association, 1977 Annual Meeting Abstracts of Papers and Symposia. New York: April 4-8, 1977, p. 60.